製品事故情報へのテキストマイニング手法の適用

Application of a text mining method to information on product related accidents

○久本誠一*,**

*製品評価技術基盤機構 事故リスク情報分析室

**京都工芸繊維大学　総合プロセーシス研究センター

HISAMOTO Seiichi*, **

*National Institute of Technology and Evaluation,

**Holistic Prosthetics Research Center of Kyoto Institute of Technology

1.　はじめに

過去に発生した製品事故の情報を整理してデータベース化し公開することは，新たに製品を設計する際に同様の事故を発生させないための検討を加えたり（未然防止），既存の類似製品により同様の事故が発生しないように対策を講じる観点から（再発防止），極めて有用である．一方，このようなデータベースでは一般に検索機能を有するが，それだけでは折角の情報を十分に活用できるとは考えにくい．特に事故の内容や事故原因，対策等の，自由記述された膨大なテキスト情報に基づいて傾向の分析や問題点の抽出を行うことは，事故発生日や製品の型式，発生場所等の，数値や簡単な単語で表現された情報の分析に比して極めて困難な作業となる．そこで本稿では，自由記述のテキスト情報を分析する手段としてテキストマイニング手法を適用し，事故情報の分析を試行し，手法の有効性を検証した結果を報告する．

なお，ここでは，製品の安全性が高まる中で残された課題となっている，誤使用・不注意を要因とする事故に注目して検討をおこなった．

2.　テキストマイニングについて

テキストマイニングは，自由記述のテキストデータを分析対象とし，そのプロセスは一般に，①テキストを語素（日本語の最小単位）に分解し，②全ての語素の相互関係を解析し，③統計処理を行う，というものである．主な分析内容としては，単語の出現頻度，単語の係り受け，コレスポンデンス分析，出現する単語の因果関係の分析等である．

3.　対象データ

分析の対象データは，NITE（製品評価技術基盤機構）がホームページで公開している製品事故情報約34000 件から，事故原因が誤使用・不注意に分類された事故 5705 件(2011 年 7 月 22 日時点のデータ)とした．当該事故情報では，属性データとして「事故発生日」「年度番号」「品目」「品名」「型式・機種」「製造・輸入事業者」「製品の使用期間」「事故通知者」「被害の種類」「原因区分コード」が，テキストデータとして「事故通知内容」「事故原因」「再発防止措置」がデータベース化され，これらに対してキーワード検索を行うことができる．「年度番号」は，事故案件を整理するための個別番号で，「品名」は製品群をカテゴライズするための名称，「被害の種類」は「1.死亡」「2.重傷」「3.軽傷」「4.拡大被害」「5.製品破損」から「6.被害なし」の 6 段階に分類され，「原因区分コード」は，「設計不良」「製造不良」「品質管理不十分」「表示または取扱説明書の不備」「経年劣化」「設置・施工不良」「修理不良」「消費者の誤使用・不注意」「偶発的事故」等に分類してコード化されている．

なお，テキストデータは，元データが Microsoft Access で整理されている都合で，255 字を上限に記述されている．

4.　方法

分析にはテキストマイニングツール「True Teller」(株式会社野村総合研究所)を用い，csv 形式に変換した事故情報データをインポートして分析した．

5.　結果と考察

5.1　属性データのクロス分析

本稿で使用したデータの性質を鳥瞰するため，まず，属性データのうち「品目名」について件数をカウントした．結果を図 1 に示す．これより，誤使用・不注意に分類できる事故では，燃焼機具に関係した事故(2578 件)と家庭用電気製品(1594 件）で全体の 73.1 [%]を占める結果となった．これは，①身の回りに家庭用電気製品が数多く存在すること，②燃焼機具には，1 年を通じて使用頻度の高い製品（ガスこんろ，ガス給湯機等）や季節によっては使用頻度の高い製品（石油ストーブ等）が含まれること，等が要因と考えられる．

さらに燃焼器具の内訳をカウントすると，ガスこんろが最多で 1075 件(燃焼器具の 41.7[%])，次いで石油ストーブが 415 件（燃焼器具の 16.1[%]）で，両者をあわせると燃焼機具の過半数を占める．

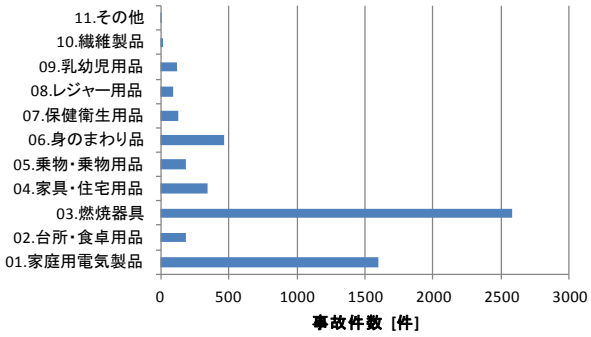

図 1　品目名と事故件数

5.2　テキストデータにおける単語の出現頻度

図 2 に，「事故内容」を対象として品詞を限定せずに単語の出現頻度をカウントした結果のうち，上位 20 位までを示す．これより，「出火する」，「全焼する」，「焼く」，等の直接火災に関連した単語や，「ガス」，「こんろ」等の，火災に関わる恐れのある製品に関連した単語の出現頻度が高いことが分かる．

また，製品使用者の誤使用・不注意に係る挙動を検討するために，「事故内容」から動詞を抽出し，出現頻度をカウントした結果を図 3 に示す．ここでも，「出火する」「全焼する」「焼損する」「焦げる」等の火災に関連した動詞の出現頻度が高いことと，併せて「過熱する」「隣接する」「置く」「転倒する」等の，使用者の挙動に係ると思われる動詞の出現が認められる．

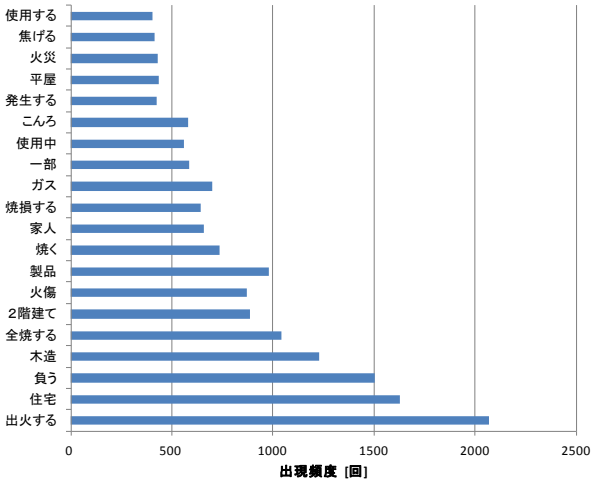

図 2　「事故内容」における単語出現頻度

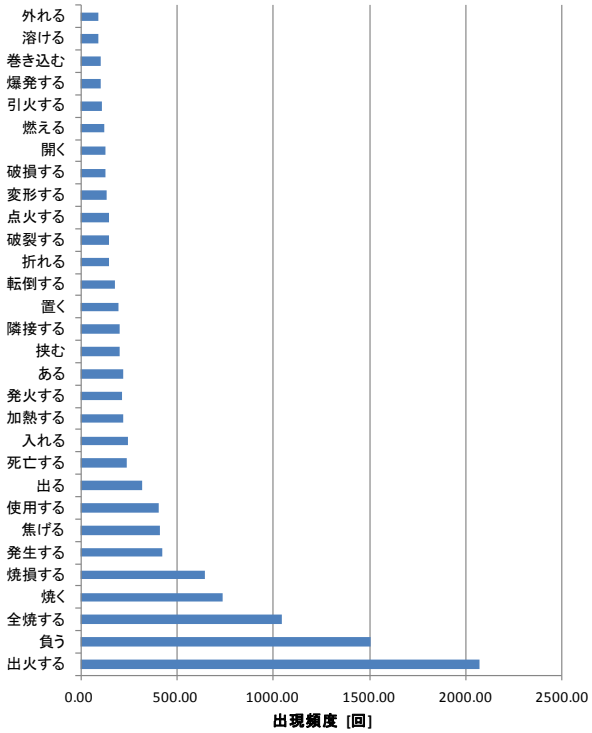

図 3　「事故内容」における動詞の出現頻度

さらに，「事故原因」から動詞を抽出して出現頻度をカウントした結果から，上位 20 位までを図 4 に示す．これより，事故原因の観点から，「過熱する」「放置する」「かける」「接触する」「誤る」等の，使用者の挙動に関連すると思われる動詞が抽出されていることがわかる．

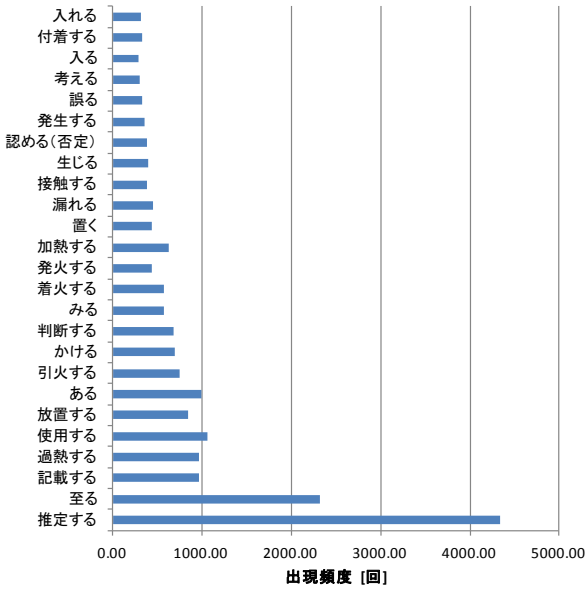

図 4　「事故原因」における動詞の出現頻度

5.3　コレスポンデンス分析

前項の結果をうけて，さらに詳細に分析するため，まず「事故内容」を対象としてコレスポンデンス分析を実施した．図 5 は，その結果を単語のプロット図（布置図）として示したもので，距離の近い単語ほどひとつのテキスト中に含まれる頻度が高い（共起性が高い）ことを示す．図 5 で表示した単語は全単語の 75[%]までで，単語相互の関連度を示す線は 100[%]を表示した．図 5 の左端の「製品」の近傍には「使用中」「火災」「上」等の単語がプロットされ，また，右端の「全焼」の近傍には「2 階建て」「木造」「住宅」等の単語がプロットされ，何れも事故発生の状況を象徴的に示している．

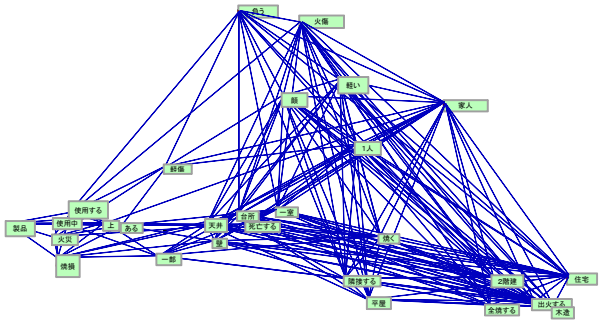

図 5　「事故内容」のコレスポンデンス分析

さらに，「事故原因」についても同様にコレスポンデンス分析を実施した結果を図 6 に示す．図 6 も全単語の 75[%]までを表示し，単語相互の関連度を示す線は 100[%]を表示している．図 6 左端の「ガス」「こんろ」の近傍に「かける」「火」がプロットされ，中央左よりには「被害者」の近傍に，「操作」「判断する」「誤る」等の単語がプロットされて，ヒューマンエラーの存在を示唆している．

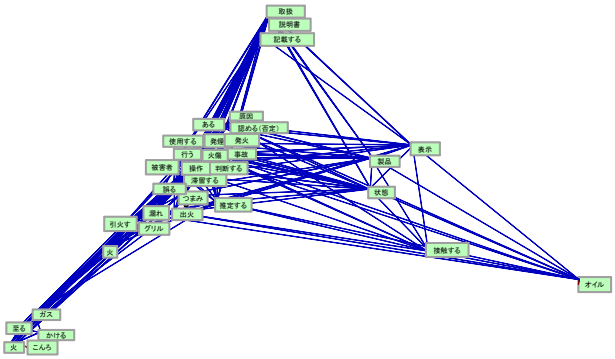

図 6　「事故原因」のコレスポンデンス分析

以上をふまえ，事故件数の多い製品としてガスこんろに係る情報を抽出し，コレスポンデンス分析を実施した．結果を図 7 に示す．図 7 の左上には，「グリル」の近傍に「油脂」「溜まる」等の単語がプロットされ，グリルの清掃不良によるグリル火災がイメージされる．図 7 の右下には，「こんろ」の近傍に「天ぷら鍋」「油」「かける」「放置する」「過熱する」等の単語がプロットされ，ガスこんろによる事故の典型例である天ぷら油火災がイメージされる．また左下には，「調理中」「ホース」「漏洩する」「火」「引火する」等の単語がプロットされ，ガス漏れがイメージされるが，その他にも多くの単語がプロットされており，複数の事故原因が分離されずにプロットされていると考えられる．

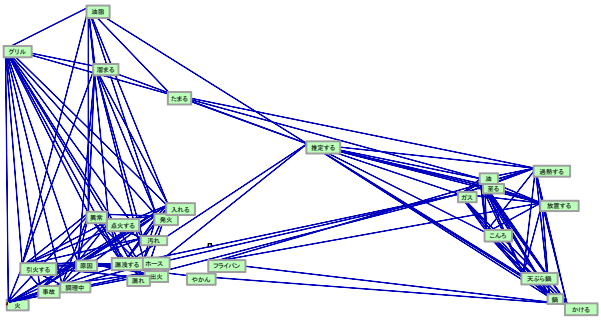

図 7　ガスこんろに係る事故の
「事故原因」のコレスポンデンス分析

5.4　因果関係分析

さらに，ガスこんろに係る事故情報について，「事故内容」と「事故原因」を対象に因果関係分析を実施した結果を図 8 に示す．ここでは事故内容と事故原因の各々について，動詞単独，並びに，名詞 – 動詞の係り受けについての分析をおこなった．図 8 では単語は抽出された全単語を表示し，因果関係を示す線は，関係の深いものから上位 30[%]を表示している．

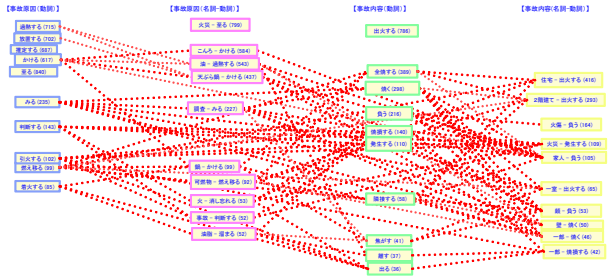

図 8　ガスこんろに係る事故の
「事故内容」と「事故原因」における因果関係分析

図 8 で, 事故原因の係り結びとしては,「こんろ – かける」(584 件),「天ぷら鍋 – かける」(437 件) の件数が多く，前者は事故内容の「家人 – 負う」(105 件) との関連が深く, 後者は事故原因の「過熱する」(715 件),「放置する」(702 件) 及び事故内容の「家人 – 負う」(105 件) との関連が深いことがわかった．これより，後者では，「天ぷら鍋をガスこんろにかけて火を付け，そのまま放置したために火災に至って，家人が火傷を負う」という，天ぷら油火災の一般的なパターンを表出していると考えられる．

6.　おわりに

テキストマイニング手法による製品事故情報分析の有効性の検証を目的として，コレスポンデンス分析及び因果関係分析を試行した．ここでは過去の事故発生件数も多く，また多く発生している事故のパターンも既知である製品として，ガスこんろを中心に解析を試行した．その結果，過去の事例にみられる，事故の典型パターンを表出する分析結果が得られた．これより，事故のパターン (モデル) が整理できていない製品 (あるいは製品群) について，テキストマイニング手法を適用すれば，事故の未然防止に有効な情報を得られると考えられる．

NITE 事故情報データベースでは，個人情報保護の観点等により，公開情報には製品使用者や被害者の属性情報がほとんど含まれておらず，分析には限界がある．しかしデータ件数が多いことから，さらに詳細な分析を行うことで，一般消費者が陥りやすい誤使用・不注意のモデル化に向けた検討を進めたい．

参考文献

1)　NITE 事故情報データベース：
http://www.jiko.nite.go.jp/

2)　川野雅央，和多田淳三，青木祐介: テキストマイニングを用いた家電製品事故情報分析, International Symposium on Management Engineering 2005, Session 5-18 (2005).

3)　早坂敏治，木村昌臣，大倉典子，土屋文人：テキストマイニング手法を用いた医薬品投薬ヒヤリハット事例の解析，第 36 回日本人間工学会関東支部大会講演集，D2-1-1，177/178 (2006)

4)　板垣晴彦：事故情報をどう活かすか，セーフティーハート，122 号(2008)

5)　久本誠一：製品安全とヒューマンエラー，生活安全ジャーナル，11 , pp. 16-19(2011)